Non parametric Bayesian priors for hidden Markov random fields

Florence Forbes, Hongliang Lu, Julyan Arbel

Bayesian Nonparametric Priors for Hidden Markov Random Fields: Application to Image Segmentation

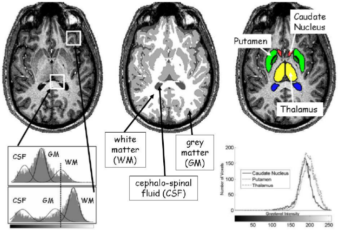
Hongliang Lü, Julyan Arbel, Florence Forbes

Inria Grenoble Rhône-Alpes and University Grenoble Alpes
Laboratoire Jean Kuntzmann
Mistis team
florence.forbes@inria.fr

July 2018

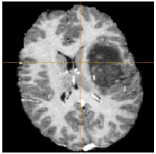
Unsupervised image segmentation

Challenges for mixture models (clustering)

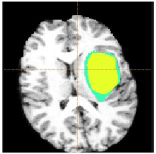


inhomogeneities, noise

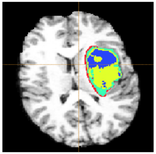
How many segments?



T1 gado



2 classes

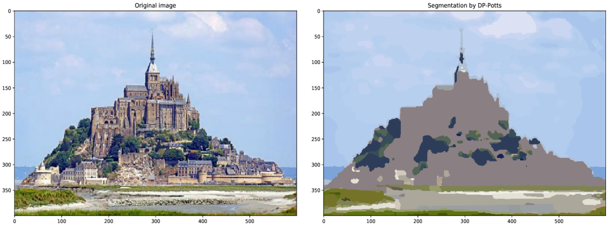


4 classes

Extensions of Dirichlet Process mixture model with spatial regularization

Outline of the talk

1. Dirichlet process (DP)
2. Spatially-constrained mixture model: DP-Potts mixture model
   - Finite mixture model
   - Bayesian finite mixture model
   - DP mixture model
   - DP-Potts mixture model
3. Inference using variational approximation
4. Some image segmentation results
5. Conclusion and future work



Dirichlet process (DP)

The DP is a central Bayesian nonparametric (BNP) prior[1].

Definition (Dirichlet process)

A Dirichlet process on the space $\mathcal{Y}$ is a **random process** G such that there exist α (concentration parameter) and G_0 (base distribution) such that for any finite partition $\{A_1, \ldots, A_p\}$ of $\mathcal{Y}$, the random vector $(P(A_1), \ldots, P(A_p))$ will be Dirichlet distributed:

$$(P(A_1), \ldots, P(A_p)) \sim \text{Dir}(\alpha G_0(A_1), \ldots, \alpha G_0(A_p))$$

Notation: $G \sim \text{DP}(\alpha, G_0)$

The DP is the infinite-dimensional generalization of the Dirichlet distribution.

[1] Ferguson, T. (1973). A Bayesian analysis of some nonparametric problems. The Annals of Statistics, 1(2):209–230.

Dirichlet process (DP) construction

A DP prior G can be constructed using three methods:

- The Blackwell-MacQueen urn scheme
- The Chinese Restaurant Process
- The Stick-Breaking construction

[2] Sethuraman, J. (1994). A constructive definition of Dirichlet priors. Statistica Sinica, 4:639-650.

Dirichlet process (DP) construction

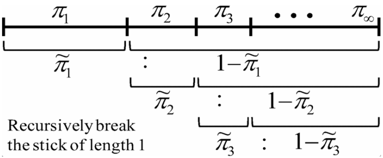
A DP prior G can be constructed using three methods:

- The Blackwell-MacQueen urn scheme
- The Chinese Restaurant Process
- The Stick-Breaking construction

The DP has almost surely discrete realizations[2]:

$$G = \sum_{k=1}^{\infty} \pi_k \delta_{\theta_k^*}$$

where $\theta_k^* \overset{\text{iid}}{\sim} G_0$ and $\pi_k = \tilde{\pi}_k \prod_{l<k}(1-\tilde{\pi}_l)$ with $\tilde{\pi}_k \overset{\text{iid}}{\sim} Beta(1,\alpha)$.



[2] Sethuraman, J. (1994). A constructive definition of Dirichlet priors. Statistica Sinica, 4:639-650.

Spatially-constrained mixture model: DP-Potts mixture

Clustering/segmentation: Finite mixture models assume data are generated by a finite sum of probability distributions:

$$\mathbf{y} = (\mathbf{y_1}, ..., \mathbf{y_N}) \text{ with } \mathbf{y_i} = (y_{i1}, ..., y_{iD}) \in \mathbb{R}^D \; i.i.d$$

$$p(\mathbf{y}_i|\theta^*, \pi) = \sum_{k=1}^{K} \pi_k F(\mathbf{y}_i|\theta_k^*)$$

where

- $\theta^* = (\theta_1^*, ..., \theta_K^*)$ and $\pi = (\pi_1, ..., \pi_K)$ with θ^* class parameters and π mixture weights with $\sum_{i=1}^{K} \pi_i = 1$.
- θ^* and π can be estimated using EM algorithm.

Equivalently

- $G = \sum_{k=1}^{K} \pi_k \delta_{\theta_k^*}$ non random
- $\theta_i \sim G$ and $\mathbf{y}_i|\theta_i \sim F(\mathbf{y}_i|\theta_i)$.

Bayesian finite mixture model

In a Bayesian setting, a prior distribution is placed over θ^* and π.

Thus, the posterior distribution of parameters given the observations is

$$p(\theta^*, \pi|\mathbf{y}) \propto p(\mathbf{y}|\theta^*, \pi)p(\theta^*, \pi)$$

To generate a data point within a **Bayesian finite mixture model**:

- $\theta_k^* \sim G_0$
- $\pi_1, ..., \pi_K \sim$ Dir$(\alpha/K, ..., \alpha/K)$
- $G = \sum_{k=1}^{K} \pi_k \delta_{\theta_k^*}$ is a random measure
- $\theta_i|G \sim G$, which means $\theta_i = \theta_k^*$ with probability π_k
- $\mathbf{y_i}|\theta_i \sim F(\mathbf{y_i}|\theta_i)$

Bayesian finite mixture model

In a Bayesian setting, a prior distribution is placed over θ^* and π.

Thus, the posterior distribution of parameters given the observations is

$$p(\theta^*, \pi | \mathbf{y}) \propto p(\mathbf{y} | \theta^*, \pi) p(\theta^*, \pi)$$

To generate a data point within a **Bayesian finite mixture model**:

- $\theta_k^* \sim G_0$
- $\pi_1, ..., \pi_K \sim$ Dir$(\alpha/K, ..., \alpha/K)$
- $G = \sum_{k=1}^{K} \pi_k \delta_{\theta_k^*}$ is a random measure
- $\theta_i | G \sim G$, which means $\theta_i = \theta_k^*$ with probability π_k
- $\mathbf{y_i} | \theta_i \sim F(\mathbf{y_i} | \theta_i)$

Limitation:

Require specifying the number of components K beforehand.

Solution:

Assume an infinite number of components using BNP priors.

DP mixture model

From a Bayesian finite mixture model to a DP mixture model

To establish a DP mixture model, let G be a DP prior ($K \to \infty$), namely

$$G \sim \mathrm{DP}(\alpha, G_0)$$

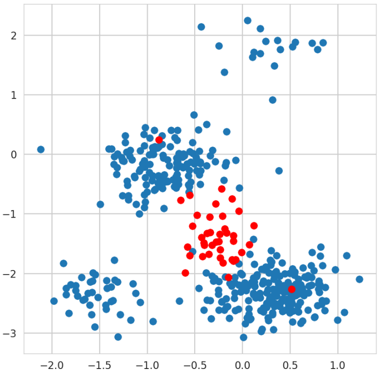
and complement it with a likelihood associated to each θ_i

To generate a data point within a **DP mixture model**:

- $G \sim \mathrm{DP}(\alpha, G_0)$
- $\theta_i | G \sim G$
- $\mathbf{y_i} | \theta_i \sim F(\mathbf{y_i} | \theta_i)$

DP mixture model

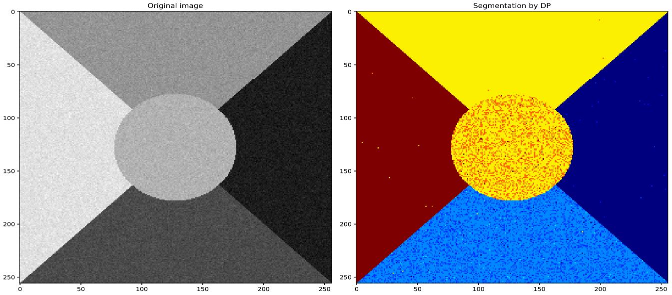
2D point clustering (unsupervised learning) based on the DP mixture model:



Let the data speak for themselves!

DP mixture model

Application to image segmentation:



Drawback:

Spatial constraints and dependencies are not considered.

Solution:

Combine the DP prior with a hidden Markov random field (HMRF).

DP-Potts mixture model

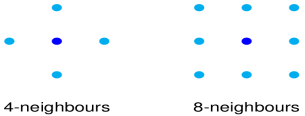
To solve the issue, we introduce a spatial Potts model component:

$$M(\theta) \propto \exp\left(\beta \sum_{i \sim j} \delta_{z(\theta_i)=z(\theta_j)}\right)$$

with $\theta = (\theta_1, ..., \theta_N)$ and β the interaction parameter.

The DP mixture model is thus extended:

- $G \sim \mathrm{DP}(\alpha, G_0)$
- $\theta | M, G \sim M(\theta) \times \prod_i G(\theta_i)$
- $\mathbf{y_i} | \theta_i \sim F(\mathbf{y_i} | \theta_i)$



DP-Potts mixture model

Other spatially-constrained BNP mixture models + inference algorithms:

- DP or PYP-Potts partition model + MCMC[3]
- Hemodynamic brain parcellation (DP-Potts) + PARTIAL VB[4]
- DP or PYP-Potts + Iterated Conditional Mode (ICM)[5]

Markov chain Monte Carlo (MCMC):

- Advantage: asymptotically exact
- Drawback: computationally expensive

Variational Bayes (VB):

- Advantage: much faster
- Drawback: less accurate, no theoretical guarantee

We propose **a DP-Potts mixture model** based on **a general stick-breaking construction** that allows **a natural Full VB algorithm** enabling scalable inference for large datasets and straightforward generalization to other priors.

[3] Orbanz & Buhmann (2008); Xu, Caron & Doucet (2016); Sodjo, Giremus, Dobigeon & Giovannelli (2017)

[4] Albughdadi, Chaari, Tourneret, Forbes, Ciuciu (2017)

[5] Chatzis & Tsechpenakis (2010); Chatzis (2013)

DP-Potts: Stick breaking construction

Stick breaking construction of DP: $G \sim DP(\alpha, G_0)$

- $\theta_k^* | G_0 \sim G_0$
- $\tau_k | \alpha \sim \mathcal{B}(1, \alpha), k = 1, 2, \ldots$
- $\pi_k(\tau) = \tau_k \prod_{l=1}^{k-1} (1 - \tau_l), k = 1, 2, \ldots$
- $G = \sum_{k=1}^{\infty} \pi_k(\tau) \delta_{\theta_k^*}$

+

- $\theta_i | G \sim G$
- $\mathbf{y_i} | \theta_i \sim F(\mathbf{y_i} | \theta_i)$

= Dirichlet Process Mixture Model (DPMM)

DP-Potts: Stick breaking construction

Stick breaking construction of DPMM

- $\theta_k^*|G_0 \sim G_0$
- $\tau_k|\alpha \sim \mathcal{B}(1,\alpha), k = 1, 2, \ldots$
- $\pi_k(\tau) = \tau_k \prod_{l=1}^{k-1}(1-\tau_l), k = 1, \ldots$
- $G = \sum_{k=1}^{\infty} \pi_k(\tau)\delta_{\theta_k^*}$ $\Longrightarrow$
- $\theta_i|G \sim G$
- $\mathbf{y_i}|\theta_i \sim F(\mathbf{y_i}|\theta_i)$

Stick breaking construction of DPMM

- $\theta_k^*|G_0 \sim G_0$
- $\tau_k|\alpha \sim \mathcal{B}(1,\alpha), k = 1, 2, \ldots$
- $\pi_k(\tau) = \tau_k \prod_{l=1}^{k-1}(1-\tau_l), k = 1, \ldots$
- $\theta_i = \theta_k^*$ with probability $\pi_k(\tau)$
- $\mathbf{y_i}|\theta_i \sim F(\mathbf{y_i}|\theta_i)$

DP-Potts: Stick breaking construction

Using assignment variables z_i

DPMM view

- $\theta_k^* | G_0 \sim G_0$
- $\tau_k | \alpha \sim \mathcal{B}(1, \alpha), k = 1, 2, \ldots$
- $\pi_k(\tau) = \tau_k \prod_{l=1}^{k-1} (1 - \tau_l), k = 1, \ldots$
- $\theta_i = \theta_k^*$ with probability $\pi_k(\tau)$
- $\mathbf{y_i} | \theta_i \sim F(\mathbf{y_i} | \theta_i)$ $\Longrightarrow$

Mixture/Clustering view

- $\theta_k^* | G_0 \sim G_0$
- $\tau_k | \alpha \sim \mathcal{B}(1, \alpha), k = 1, 2, \ldots$
- $\pi_k(\tau) = \tau_k \prod_{l=1}^{k-1} (1 - \tau_l), k = 1, \ldots$
- $p(z_i = k | \tau) = \pi_k(\tau)$
- with $z_i = z(\theta_i) = k$ when $\theta_i = \theta_k^*$
- $\mathbf{y_i} | z_i, \theta^* \sim F(\mathbf{y_i} | \theta_{z_i}^*)$

DP-Potts: Stick breaking construction

Using assignment variables z_i

Stick breaking of DPMM

- $\theta_k^*|G_0 \sim G_0$
- $\tau_k|\alpha \sim \mathcal{B}(1,\alpha), k = 1, 2, \ldots$
- $\pi_k(\tau) = \tau_k \prod_{l=1}^{k-1}(1-\tau_l)$
- $p(z_i = k|\tau) = \pi_k(\tau)$
- $\mathbf{y_i}|z_i, \theta^* \sim F(\mathbf{y_i}|\theta^*_{z_i})$

Stick breaking of DP-Potts

- $\theta_k^*|G_0 \sim G_0$
- $\tau_k|\alpha \sim \mathcal{B}(1,\alpha), k = 1, 2, \ldots$
- $\pi_k(\tau) = \tau_k \prod_{l=1}^{k-1}(1-\tau_l)$
- $p(\mathbf{z}|\tau,\beta) \propto \prod_i \pi_{z_i}(\tau) \exp(\beta \sum_{i\sim j} \delta_{z_i=z_j})$

  $\mathbf{z} = \{z_1, \ldots, z_N\}$
- $\mathbf{y_i}|z_i, \theta^* \sim F(\mathbf{y_i}|\theta^*_{z_i})$

NB: Well defined for every stick breaking construction ($\sum_{k=1}^{\infty} \pi_k = 1$) :

e.g. Pitman-Yor $(\tau_k|\alpha,\sigma) \sim \mathcal{B}(1-\sigma, \alpha + k\sigma)$

Inference using variational approximation

Clustering/ segmentation task:

- Estimating $\mathbf{Z}$
- while parameters $\boldsymbol{\Theta}$ unknown , eg. $\boldsymbol{\Theta} = \{\boldsymbol{\tau}, \alpha, \boldsymbol{\theta}^*\}$

Bayesian setting

Access the intractable $p(\mathbf{Z}, \boldsymbol{\Theta}|\mathbf{y}, \boldsymbol{\Phi})$ approximate as $q(\mathbf{z}, \boldsymbol{\Theta}) = q_z(\mathbf{z})q_\theta(\boldsymbol{\Theta})$

Variational Expectation-Maximization

Alternate maximization in q_z and q_θ (ϕ are hyperparameters) of the Free Energy:

$$\begin{aligned}\mathcal{F}(q_z, q_\theta, \phi) &= E_{q_z q_\theta}\left[\log \frac{p(\mathbf{y}, \mathbf{Z}, \boldsymbol{\Theta}|\phi)}{q_z(\mathbf{z})q_\theta(\boldsymbol{\Theta})}\right] \\ &= \log p(\mathbf{y}|\phi) - KL(q_z q_\theta, p(\mathbf{Z}, \boldsymbol{\Theta}|\mathbf{y}, \phi))\end{aligned}$$

DP-Potts Variational EM procedure

Joint DP-Potts (Gaussian) Mixture distribution

$$p(\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{\tau}, \alpha, \boldsymbol{\theta}^* | \boldsymbol{\phi}) = \prod_{j=1}^{N} p(y_j | z_j, \boldsymbol{\theta}^*) \; p(\boldsymbol{z} | \boldsymbol{\tau}, \beta) \; \prod_{k=1}^{\infty} p(\tau_k | \alpha) \prod_{k=1}^{\infty} p(\theta_k^* | \rho_k) \; p(\alpha | s_1, s_2)$$

- $p(y_j | z_j, \boldsymbol{\theta}^*) = \mathcal{N}(y_j | \mu_{z_j}, \Sigma_{z_j})$ is Gaussian
- $p(\boldsymbol{z} | \boldsymbol{\tau}, \beta)$ is a DP-Potts model
- $p(\tau_k | \alpha)$ is Beta $\mathcal{B}(1, \alpha)$
- $p(\theta_k^* | \rho_k) = \mathcal{NIW}(\mu_k, \Sigma_k | m_k, \lambda_k, \Psi_k, \nu_k)$ is Normal-inverse-Wishart
- $p(\alpha | s_1, s_2) = \mathcal{G}(\alpha | s_1, s_2)$ is Gamma

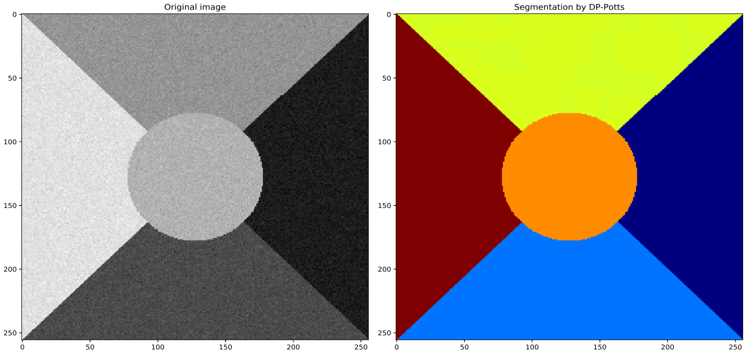
Usual truncated variational posterior, $q_{\tau_k} = \delta_1$ for $k \geq K$ (eg. $K = 40$)

$$q(\mathbf{z}, \boldsymbol{\Theta}) = \prod_{j=1}^{N} q_{z_j}(z_j) \; q_\alpha(\alpha) \prod_{k=1}^{K-1} q_{\tau_k}(\tau_k) \prod_{k=1}^{K} q_{\theta_k^*}(\mu_k, \Sigma_k)$$

- E-steps: VE-Z, VE-α, VE-$\boldsymbol{\tau}$ and VE-$\boldsymbol{\theta}^*$
- M-step: ϕ updating straightforward except for β

Some image segmentation results

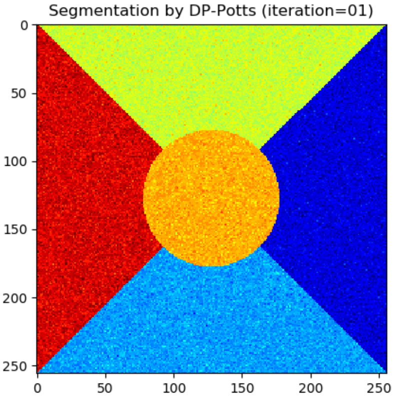
Model validation and verification:



Segmented image using DP-Potts model with $\beta = 2.5$.

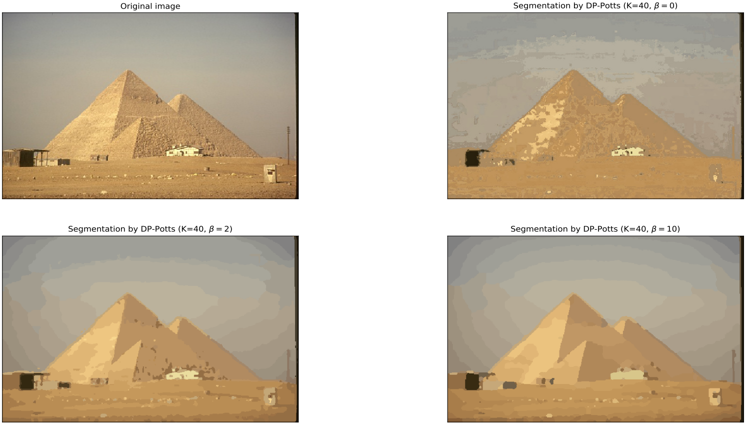
Some image segmentation results

Convergence of the VB algorithm initialized by the k-means++ clustering:



Some image segmentation results

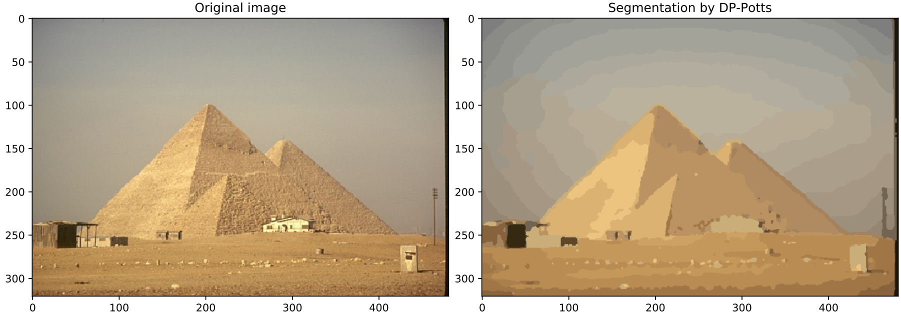
Segmentation results for Berkeley Segmentation Dataset:



The segmentation results obtained by DP-Potts model with $\beta = 0, 1, 5$.

Some image segmentation results

Segmentation with estimated $\beta = 1.66$



Quantitative evaluation of the segmentations

Probabilistic Rand Index on 154 color (RGB) images with ground truth (several) from Berkeley dataset (1000 superpixels). But Manual ground truth segmentations are subjective !

PRI results with DP-Potts model

	Mean	Median	St.D.
K=10	71.48	72.54	0.1040
K=20	73.64	73.42	0.0935
K=40	75.33	76.47	0.0853
K=50	75.81	76.31	0.0873
K=60	76.55	77.12	0.0848
K=80	**77.06**	**78.30**	**0.0835**

PRI results from Chatzis 2013

PRI (%)	DPM	iHMRF	MRF-PYP	GC
Mean	74.15	75.50	76.49	76.10
Median	75.49	76.89	78.08	77.59
St.D.	0.084	0.082	0.079	0.083

Computation time : Berkeley 321x481 image reduced to 1000 superpixels takes **10-30 s** on a PC with CPU Intel(R) Core(TM) i7-5500U CPU 2.40GHz and 8GB RAM

Conclusion and future work

- A general DP-Potts model and the associated VB algorithm were built.
- The DP-Potts model was applied to image segmentation and tested on different types of datasets.
- Impact of the interaction parameter β on the final results is significant.
- An estimation procedure was proposed for β

Conclusion and future work

- A general DP-Potts model and the associated VB algorithm were built.
- The DP-Potts model was applied to image segmentation and tested on different types of datasets.
- Impact of the interaction parameter β on the final results is significant.
- An estimation procedure was proposed for β

- How does β influence the number of components?
- Extend the model with other priors (Pitman-Yor process, Gibbs-type priors, etc.).
- Try other variational approximations (truncation-free)
- Investigate theoretical properties of BNP priors under structural constraints (time, spatial) ....
- ... for other applications, such as discovery probabilities, etc.

References

1. M. Albughdadi, L. Chaari, J.-Y. Tourneret, F. Forbes, and P. Ciuciu. A Bayesian non-parametric hidden Markov random model for hemodynamic brain parcellation. Signal Processing, 135:132-146, 2017.
2. S. P. Chatzis. A Markov random field-regulated Pitman-Yor process prior for spatially constrained data clustering. Pattern Recognition, 46(6): 1595-1603, 2013.
3. S. P. Chatzis and G. Tsechpenakis. The infinite hidden Markov random field model. IEEE Trans. Neural Networks, 21(6):1004-1014, 2010.
4. F. Forbes and N. Peyrard. Hidden Markov Random Field Selection Criteria based on Mean Field-like approximations. IEEE PAMI, 25(9):1089-1101, 2003
5. P. Orbanz and J. M. Buhmann. Nonparametric Bayesian image segmentation. International Journal of Computer Vision, 77(1-3):25-45, 2008.
6. J. Sodjo, A. Giremus, N. Dobigeon, J.F. Giovannelli, A generalized Swendsen-Wang algorithm for Bayesian nonparametric joint segmentation of multiple images, ICASSP, 2017.
7. R. Xu, F. Caron, and A. Doucet. Bayesian nonparametric image segmentation using a generalized Swendsen-Wang algorithm. ArXiv e-prints, February 2016.

Thank you for your attention!

contact: florence.forbes@inria.fr

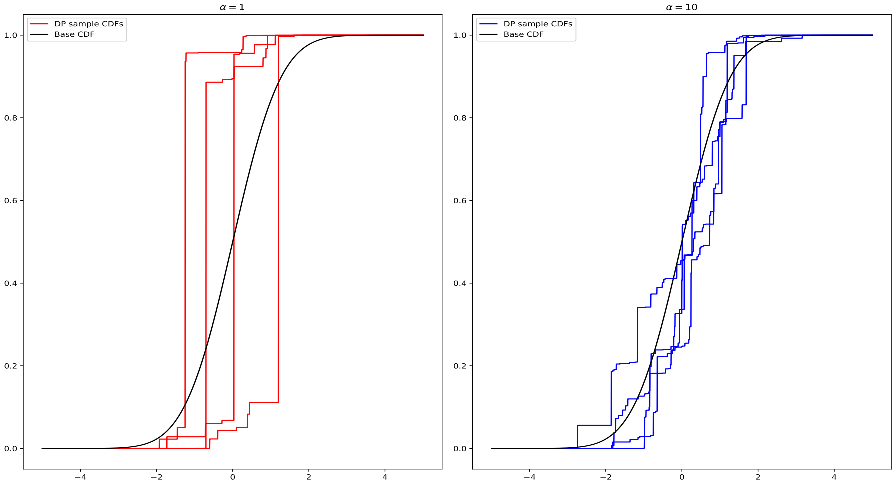
Job opportunity

Data Institute
Univ. Grenoble Alpes

Université Grenoble Alpes invites applications for a **2-year junior research chair** (post-doc) in Data Science for Life Sciences and Health

- Starting in October 2018
- Data science methodology and machine learning to Life Sciences and Health
- Application deadline: August, 31 2018
- Website: https://data-institute.univ-grenoble-alpes.fr/
- Contact: florence.forbes@inria.fr

Stick breaking construction



DP simulations with G_0 being a standard normal distribution $\mathcal{N}(0, 1)$ and $\alpha = 1, 10$ using the Stick-Breaking representation.

Variational EM

General formulation, at iteration (r)

E-Z $q_z^{(r)}(\mathbf{z}) \propto \exp\left(E_{q_\theta^{(r-1)}}[\log p(\mathbf{y}, \mathbf{z}, \boldsymbol{\Theta}|\boldsymbol{\phi}^{(r-1)})]\right)$

E-Θ $q_\theta^{(r)}(\boldsymbol{\Theta}) \propto \exp\left(E_{q_z^{(r)}}[\log p(\mathbf{y}, \mathbf{Z}, \boldsymbol{\Theta}|\boldsymbol{\phi}^{(r-1)})]\right)$

M-ϕ $\boldsymbol{\phi}^{(r)} = arg\max_{\boldsymbol{\phi}} E_{q_z^{(r)} q_\theta^{(r)}}[\log p(\mathbf{y}, \mathbf{Z}, \boldsymbol{\Theta}|\boldsymbol{\phi})]$

VE-Z, VE-α, VE-$\boldsymbol{\tau}$, and VE-$\boldsymbol{\theta}^*$

e.g. VE-Z step divides into N VE-Z_j steps ($q_{z_j}(z_j) = 0$ for $z_j > K$)

$$q_{z_j}(z_j) \propto \exp\left(\mathrm{E}_{q_{\theta^*_{z_j}}}\left[\log p(y_j|\theta^*_{z_j})\right] + \mathrm{E}_{q_\tau}\left[\log \pi_{z_j}(\boldsymbol{\tau})\right] + \beta \sum_{i \sim j} q_{z_i}(z_j)\right)$$

Estimation of β

M-β step: involves $p(\boldsymbol{z}|\boldsymbol{\tau},\beta) = \mathcal{K}(\beta,\boldsymbol{\tau})^{-1}\exp(V(\boldsymbol{z};\boldsymbol{\tau},\beta))$

with $V(\boldsymbol{z};\boldsymbol{\tau},\beta) = \sum_i \log \pi_{z_i}(\boldsymbol{\tau}) + \beta \sum_{i\sim j} \delta_{(z_i=z_j)}$

$$\begin{aligned}\hat{\beta} &= \arg\max_{\beta} \mathrm{E}_{q_z q_\tau}\left[\log p(\boldsymbol{z}|\boldsymbol{\tau};\beta)\right] \\ &= \arg\max_{\beta} \mathrm{E}_{q_z q_\tau}\left[V(\boldsymbol{z};\boldsymbol{\tau},\beta)\right] - \mathrm{E}_{q_\tau}\left[\log \mathcal{K}(\beta,\boldsymbol{\tau})\right]\end{aligned}$$

Two difficulties

(1) $p(\boldsymbol{z}|\boldsymbol{\tau},\beta)$ is intractable (normalizing constant $\mathcal{K}(\beta,\boldsymbol{\tau})$, typical of MRF)

(2) it depends on $\boldsymbol{\tau}$ (typical of DP)

Two approximations

(1) "standard" Mean Field like approximation[a]

(2) Replace the random $\boldsymbol{\tau}$ by a fixed $\tilde{\boldsymbol{\tau}} = E_{q_\tau}[\boldsymbol{\tau}]$

[a] Forbes & Peyrard 2003

Approximation of $p(\boldsymbol{z}|\boldsymbol{\tau};\beta)$

$$p(\boldsymbol{z}|\boldsymbol{\tau},\beta) \approx \tilde{q}_z(\boldsymbol{z}|\beta) = \prod_{j=1}^{N} \tilde{q}_{z_j}(z_j|\beta)$$

$$\tilde{q}_{z_j}(z_j = k|\beta) = \frac{\exp(\log \pi_k(\tilde{\boldsymbol{\tau}}) + \beta \sum_{i\in N(j)} q_{z_i}(k))}{\sum_{l=1}^{\infty} \exp(\log \pi_l(\tilde{\boldsymbol{\tau}}) + \beta \sum_{i\in N(j)} q_{z_i}(l))} \quad \text{and} \quad \tilde{\boldsymbol{\tau}} = E_{q_\tau}[\boldsymbol{\tau}]$$

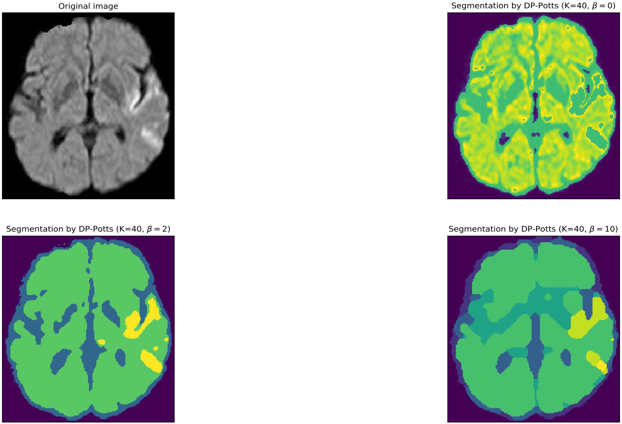
β is estimated at each iteration by setting the approximate gradient to 0

$$\mathrm{E}_{q_z q_\tau}\left[\nabla_\beta V(\boldsymbol{z};\boldsymbol{\tau},\beta)\right] = \sum_{k=1}^{K}\sum_{i\sim j} q_{z_j}(k)\; q_{z_i}(k)$$

$$\nabla_\beta \mathrm{E}_{q_\tau}\left[\log \mathcal{K}(\beta,\boldsymbol{\tau})\right] = \mathrm{E}_{p(\boldsymbol{z}|\boldsymbol{\tau},\beta)q_\tau}\left[\nabla_\beta V(\boldsymbol{z};\boldsymbol{\tau},\beta)\right] \approx \sum_{k=1}^{K}\sum_{i\sim j} \tilde{q}_{z_j}(k|\beta)\; \tilde{q}_{z_i}(k|\beta)$$

Some image segmentation results

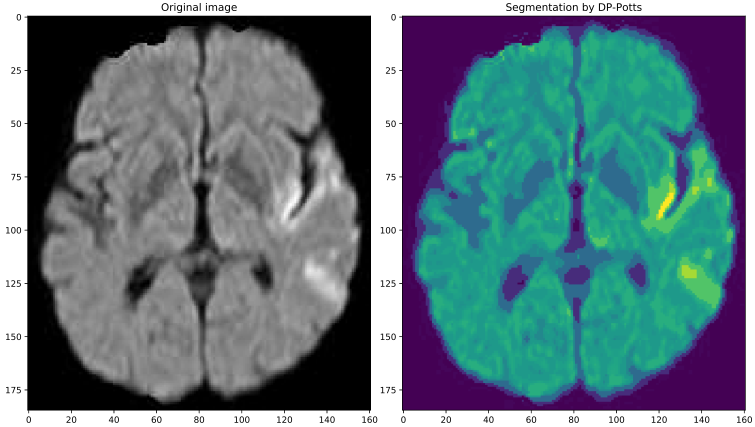
Segmentation results for medical images: all hyperparameters fixed



The segmentation results obtained by DP-Potts model with $\beta = 0, 1, 5$.

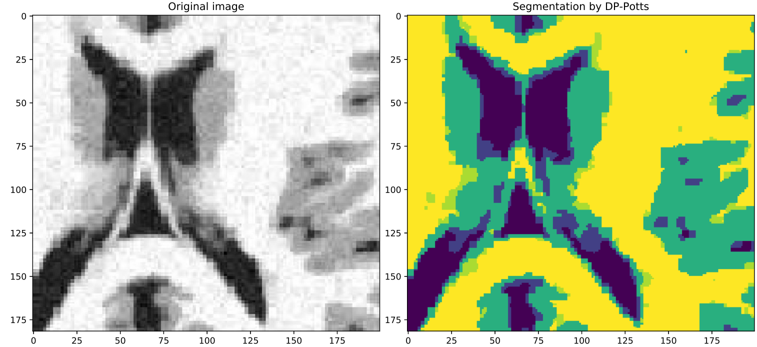
Some image segmentation results

Segmentation with estimated hyperparameters ($\beta = 0.75$)



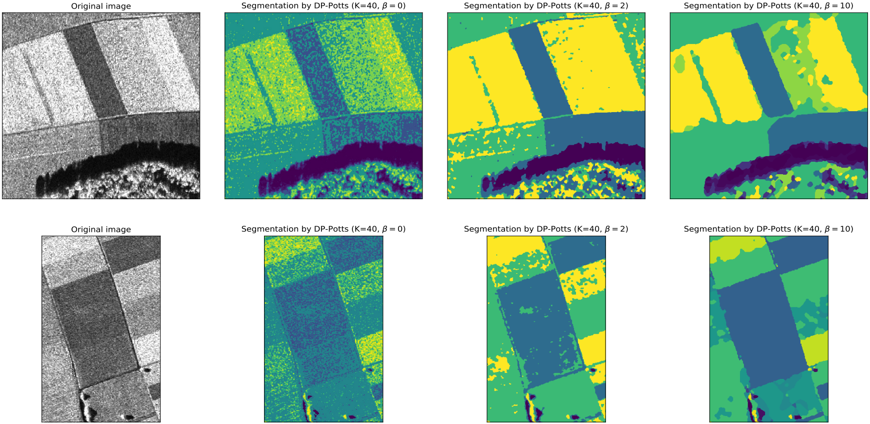
Some image segmentation results

Segmentation with estimated $\beta = 0.96$ (pixels with partial volume)



Some image segmentation results

Segmentation results for SAR images:

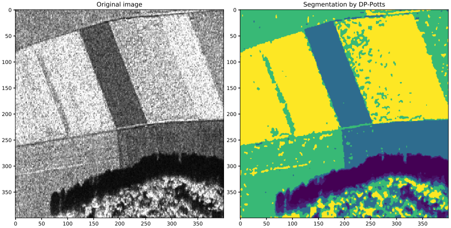


The segmentation results obtained by DP-Potts model with $\beta = 0, 1, 5$.

Some image segmentation results

Segmentation results with estimated β

$\beta = 1.11$



$\beta = 1.02$